

Pràctica 4. Manipulació d'un arxiu de dades

En aquesta pràctica veurem com podem manipular un arxiu de dades. Començarem veient com es fa per afegir o suprimir valors i variables a la nostra base de dades. A continuació veurem com es pot reordenar l'arxiu de dades segons els valors que pren una determinada variable. També veurem com podem seleccionar els individus que satisfan una determinada característica (o determinades característiques) i finalment veurem com podem segmentar un arxiu de dades, és a dir, obtenir els resultats per separat per cada un dels diferents grups definits per una variable.

► Recordeu activar les opcions d'edició:

- “Mostrar comandos en anotaciones” a la pestanya de “Visor”.
- “Nombre y etiquetas” per a les variables i “Valores y etiquetas” per als valors a l'apartat de “Etiquetado de tablas pivote” de la pestanya de “Etiquetas de resultados”.

► En aquesta pràctica treballarem amb els arxius **GSS93 reducido.sav** i **Mundo 95.sav** del programari, ubicats tots dos a la carpeta de l'SPSS.

1. Afegir individus o variables a la base de dades

A la **pràctica 1** ja vam veure com podem eliminar un individu o una variable de la base de dades. Recordarem a continuació les diferents formes de fer-ho.

• Afegir un individu

Per afegir un individu a la nostra base de dades cal estar situat a la finestra de *Vista de dades*. Un cop situats aquí

- seleccionem primer una fila per a determinar el punt d'inserció i, a continuació,
- a través del menú contextual cliquem sobre “Insertar caso”. Aquesta acció es pot executar també a través del menú “Datos” o bé a través de la icona de la barra d'eines:



• Afegir una variable

Per afegir una variable a la nostra base de dades podem estar situats a la finestra de *Vista de dades* o bé a la finestra *Vista de variables*. Si estem a la finestra *Vista de variables*,

- seleccionem primer una fila per a determinar el punt d'inserció i, a continuació,
- a través del menú contextual cliquem sobre “Insertar variables”. Aquesta acció es pot executar també a través del menú “Datos” o bé a través de la icona de la barra d'eines:



Si estem situats a la finestra *Vista de dades* cal seleccionar la columna on afegirem la variable i continuar igual que en el cas anterior.

► Exercici 1

Comproveu el funcionament de totes aquestes accions amb l'arxiu **GSS93 reducido.sav**. És a dir, comproveu com podríeu afegir individus a la base de dades i com podríeu afegir-hi variables.

2. Suprimir individus o variables a la base de dades

Recordarem a continuació les diferents formes d'eliminar files i columnes.

• Suprimir un individu

Per eliminar un individu de la nostra base de dades cal estar situat a la finestra de *Vista de dades*. Un cop situats aquí eliminar un individu equival a eliminar una fila d'aquesta matriu de dades. Per fer-ho cal:

- seleccionar la fila (o files) que volem suprimir, i
- prémer sobre <supr> o fer Ctrl+X o anar al menú “Edición” i seleccionar “Eliminar” o escollim “Borrar” en el menú contextual.

• Suprimir una variable

Per eliminar una variable de la base de dades podem fer-ho des la finestra de *Vista de dades* o bé des de la finestra *Vista de variables*. Si ho fem des de la finestra *Vista de variables*, eliminar una variable equival a eliminar una fila d'aquesta matriu de dades. Per fer-ho cal procedir igual com quan volíem eliminar un individu, és a dir,

- seleccionem la fila (o files) que volem suprimir, i
- premem sobre <supr> o fem Ctrl+X o anem al menú “Edición” i seleccionem “Eliminar” o escollim “Borrar” en el menú contextual.

Si ho fem des de la finestra de *Vista de dades*, eliminar una variable equival a eliminar una columna d'aquesta matriu de dades. Així doncs, en aquest cas:

- seleccionem la columna (o columnes) que volem suprimir, i
- premem sobre <supr> o fem Ctrl+X o anem al menú “Edición” i seleccionem “Eliminar” o escollim “Borrar” en el menú contextual.

► Exercici 2

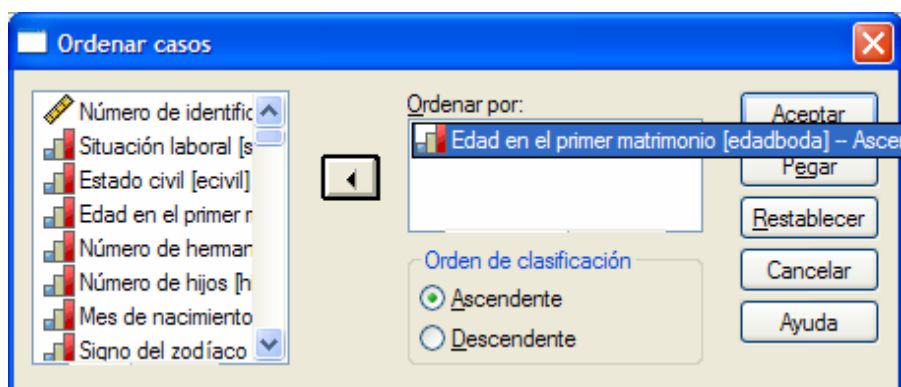
Obriu l'arxiu **GSS93 reducido.sav** i elimineu els individus 7,8 i 11 i la variable **clásica**. És important que tingueu en compte que si ara guardeu l'arxiu perdreu la informació que heu suprimit.

3. Reordenar la base de dades

A continuació veurem com podem reordenar la nostra base de dades. Per fer-ho anem al menú:

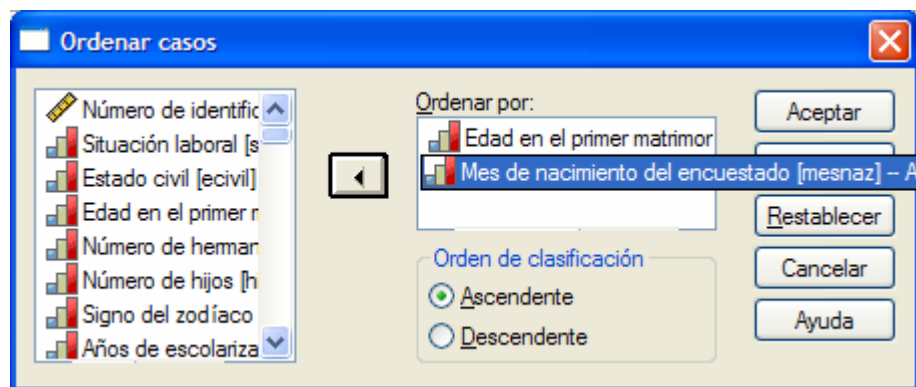
Datos / Ordenar casos

I escollim la variable, o variables, a partir dels valors de les quals volem reordenar les dades. Podem, per exemple, a l'arxiu **GSS93 reducido.sav** ordenar les dades segons la variable **edad** en ordre *ascendent*:



Observeu l'errada: escriu la variable *edad* boda en el lloc d'*edad* , però ordena per l'*edat* !!

Observeu els canvis en l'arxiu de dades. Si volem afinar més podem posar a més a més de la variable **edad** (en ordre *ascendent*) la variable **mesnaz** (en ordre *descendent*) per a que ens ordeni els individus que tenen la mateixa edat segons el seu mes de naixement, d'aquesta manera entre els individus d'un mateix any apareixen primer els més joves:



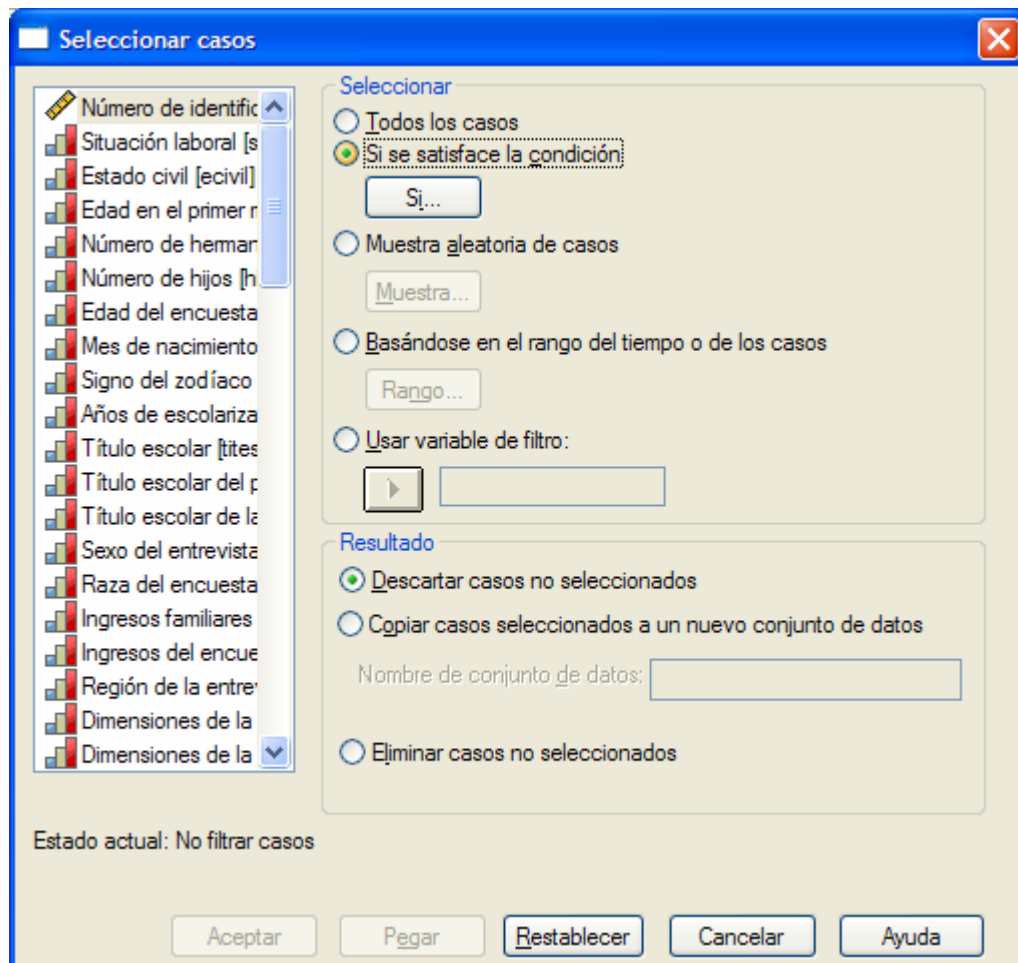
4. Seleccionar individuos

Sovint, quan treballem amb una base de dades ens interessa obtenir informació sobre els individus que satisfan determinades condicions. Ens pot interessar, per exemple, estudiar diverses variables però només per als individus de sexe femení, o per als individus que tenen fills, o per als individus de sexe masculí i amb uns ingressos anuals inferiors als 9 000 euros.

L'SPSS ens permet seleccionar els individus que satisfan una determinada característica de forma que, a partir d'aquell moment i mentre no desfem la selecció, tots els procediments que apliquem faran referència només als individus seleccionats.

A l'arxiu **GSS93 reducido.sav** anem a seleccionar només els individus de sexe femení. Per fer-ho utilitzem el menú:

Datos / Seleccionar casos / Si se satisface la condición / Si



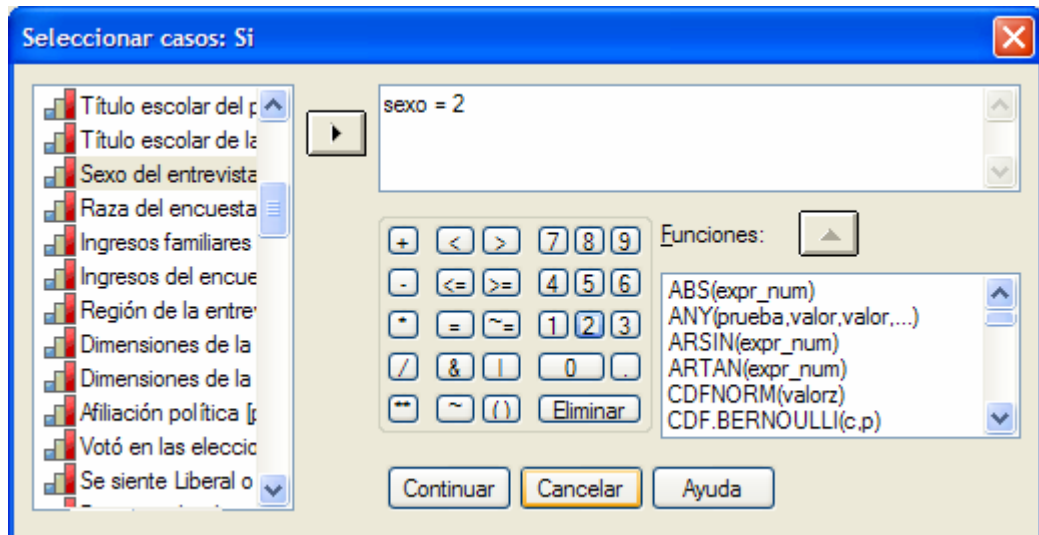
A més de l'opció "Si..." aquest procediment possibilita l'extracció d'una mostra aleatòria, la selecció a partir d'un rang de casos i fer servir variables filtre.

En qualsevol d'aquests casos podem optar per:

- **"Descartar casos no seleccionados"**: la selecció implica que les dades són filtrades, és a dir, els casos filtrats romanen a l'arxiu però s'exclouen de l'anàlisi. Per defecte, treballarem d'aquesta manera.
- **"Copiar casos seleccionados a un nuevo conjunto de datos"**, que deixaria igual l'arxiu inicial; per treballar amb els casos seleccionats caldria fer-ho des del nou arxiu.

- “Eliminar casos no seleccionados”: s’eliminen els casos no seleccionats del fitxer actiu (el de la memòria temporal del sistema). El fitxer original es manté al disc, però si després de fer la selecció dessem el fitxer amb el mateix nom llavors perdrem definitivament el casos no seleccionats. Aquesta opció és perillosa.

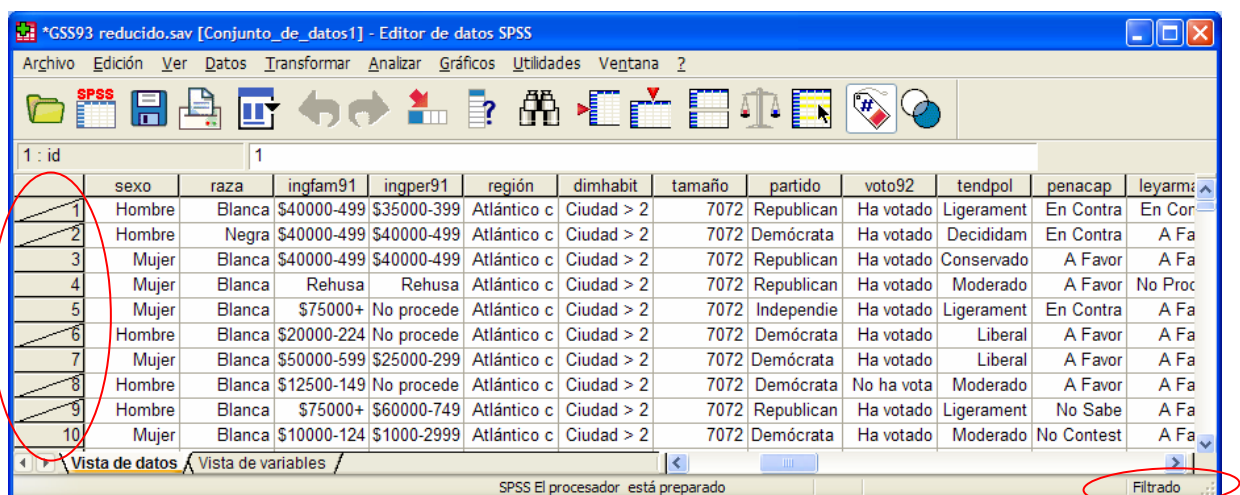
Un cop premem sobre l’icona de “Si...”, escollim la variable **sexo** i volem seleccionar només els individus per als quals aquesta variable pren el valor **2** (codificació de les dones):



Premem la tecla *Continuar* i, abans de dir *Aceptar*, comprovarem rem que, per defecte, tenim escollida l’opció

Descartar casos no seleccionados.

Si observem ara la base de dades, veurem que apareixen alguns casos tatxats al marge esquerre: són els casos que no han estat seleccionats, en el nostre cas, són els individus homes.



Observeu també que s'ha creat automàticament una nova variable de filtre anomenada **filter_\$** que pren els valors **0** i **1** amb etiquetes “No seleccionado” i “Seleccionado”, respectivament, segons si l'individu ha estat seleccionat o no. Podem canviar el nom de la variable de filtre per **filtre-sexe**, si volem mantenir-la a la base de dades; en cas contrari, quan fem un nou filtrat perdem el filtre anterior.

Observeu també que a la part inferior dreta de la finestra de l'SPSS apareix una etiqueta amb la inscripció “**Filtrado**”. Ens recorda que l'arxiu de dades amb el que treballem ha estat filtrat, és a dir, ens recorda que no estem treballant amb totes les dades sinó només amb les que satisfan una determinada característica. També ens ha aparegut a l'arxiu de resultats l'anotació de les comandes de sintaxi indicant que s'han filtrat els casos.

Si ara calculem, per exemple, la taula de freqüències de la variable **legdroga** (legalització de la marihuana) la informació obtinguda fa referència només a l'opinió de les dones de la nostra base de dades.

Legalización de la Marihuana

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	A Favor	104	12,1	18,9	18,9
	En Contra	445	51,8	81,1	100,0
	Total	549	63,9	100,0	
Perdidos	No Procede	277	32,2		
	No Sabe	32	3,7		
	No Contesta	1	,1		
	Total	310	36,1		
Total		859	100,0		

És **molt important** que, un cop haguem realitzat l'estudi que volíem fer amb només una part dels individus, ens recordem de **desfer la selecció** per tornar a treballar amb l'arxiu complet. Si no ho féssim estaríem obtenint informacions errònies. Per fer-ho, el menú és

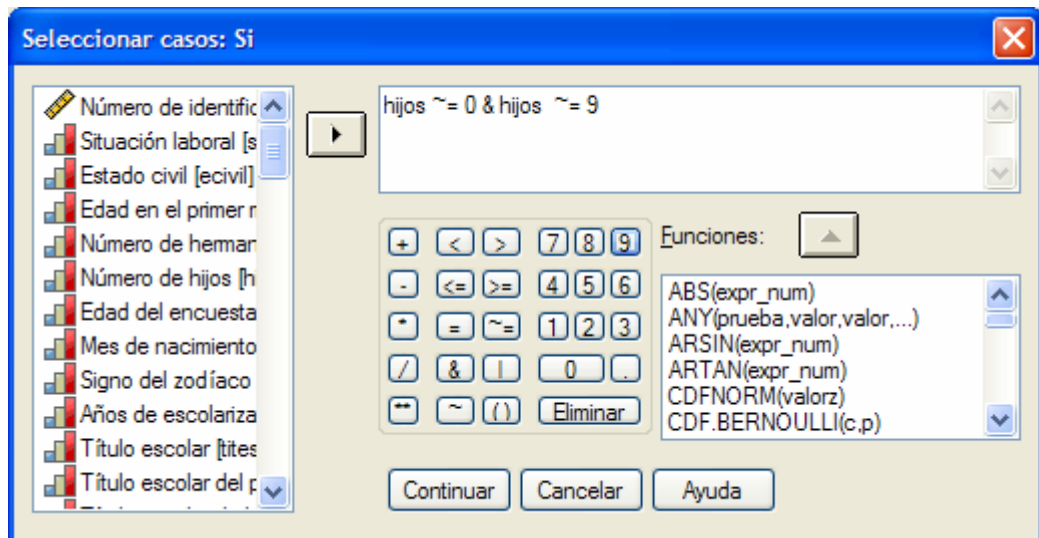
Datos / Seleccionar casos / Todos los casos / Aceptar

Veurem a continuació que també podem fer seleccions una mica més complexes.

Dins de l'arxiu amb el que estem treballant, observeu la variable **hijos** (número de fills). Suposem que volem analitzar alguna variable per als individus de la nostra mostra que són pares. Hem de descartar per tant els que no tenen cap fill (valor **0** de la variable) i els valors perduts, és a dir, els que no han respost aquesta pregunta (si observeu les etiquetes de la variable, estan codificats amb el valor **9** perquè el valor **8** ja conté tots els que tenen 8 fills o més). Així doncs hem de seleccionar tots els individus pels quals la variable **hijos** pren un valor diferent de **0** i de **9**. Per fer-ho, novament, utilitzem el menú:

Datos / Seleccionar casos / Si se satisface la condición / Si...

La condició que hem de posar és que la variable prengui un valor diferent (el símbol que cal utilitzar és $\sim =$ que significa *no igual*) de **0** i de **9**:

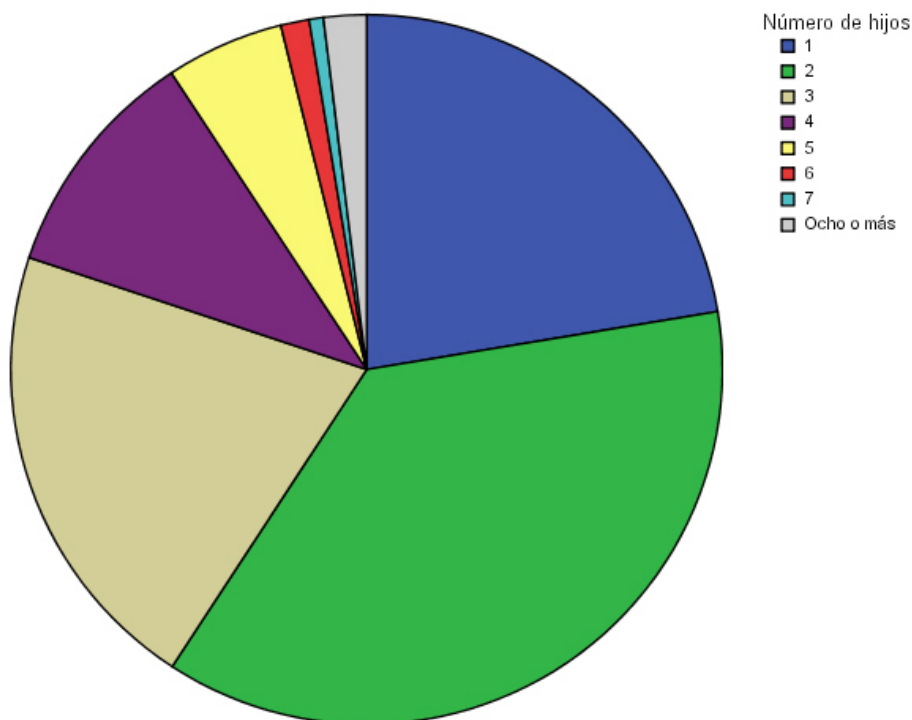


A continuació premem *Continuar* i ens assegurem que tinguem activada l'opció

Descartar casos no seleccionados

abans de prémer *Aceptar*.

Si, per exemple, fem ara un diagrama de sectors de la variable **hijos** només ens apareixeran els individus que com a mínim tenen un fill:



Abans de continuar, recordeu-vos de **desfer la selecció**.

Els principals operadors de selecció són:

= ~= < <= > >= & (AND) | (OR)

► Exercici 3

Obriu ara l'arxiu **Mundo 95.sav**. En aquest arxiu els individus són països i, per a cada país, s'ha avaluat diverses variables.

- Seleccionau els països de religió musulmana. Quan una variable és de tipus cadena, per fer la selecció cal escriure exactament el nom del valor tal i com s'ha introduït a la base de dades i posar-lo entre cometes:

"Musulma."

Un cop feta la selecció, feu un diagrama de barres de la variable region.

- Feu un diagrama de sectors de la variable region per als països que no siguin ni d'Àfrica ni d'Amèrica Llatina. (És a dir, prèviament, heu de seleccionar els països que no corresponguin a aquestes dues regions econòmiques).

- Feu ara un diagrama de sectors de la variable region per als països que tinguin un producte interior brut per càpita (variable pib_cap) inferior o igual a 10 000.

- Torneu-ho a fer però només considerant els països amb un producte interior brut per càpita entre 1 000 i 15 000.

- Repetiu-ho ara per als països que tinguin un producte interior brut per càpita inferior a 2 000 o superior a 10 000. (Podeu utilitzar la funció lògica OR).

- Finalment, seleccioneu els països que són d'Àfrica i tenen un pib_cap més petit o igual que 2000. (Podeu utilitzar la funció lògica AND o bé el símbol &). Seguidament feu un diagrama de sectors que reflecteixi la distribució de la relig (religió majoritària) dels països seleccionats

Al final, recordeu-vos de **desfer la selecció**.

5. Segmentar l'arxiu de dades

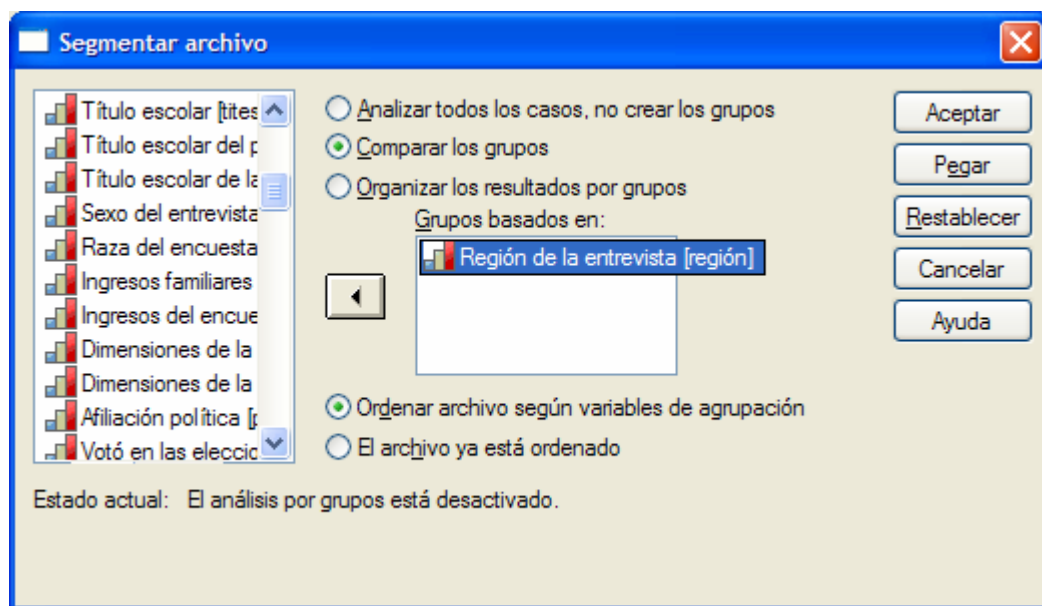
Imaginem ara que volem estudiar el producte interior brut per càpita però de forma separada per a cada una de les diferents regions econòmiques. Hem de segmentar l'arxiu de dades. És a dir, volem dividir la nostra mostra en 6 submostres (els 6 valors diferents que pot prendre la variable **region**) i fer l'estudi per a cada una d'aquestes submostres. Observeu que això ens permetrà observar si hi ha diferències de producte interior brut i quina és la magnitud d'aquestes diferències entre les diferents regions.

L'SPSS ens ofereix dues formes diferents de segmentar l'arxiu. Les veurem totes dues per tal de comparar els resultats que s'obtenen en un i altre cas.

- **Comparar els grups** El menú, en aquest cas, és el següent:

Datos / Segmentar archivo / Comparar los grupos

escollim la variable **region** i premem acceptar:



A la finestra de dades de l'SPSS només apreciem un petit canvi: a la part inferior dreta apareix l'etiqueta "**Segmentado**".

Demaneu ara que ens calculi algunes característiques de la variable **pib_cap**, per exemple, la mitjana, la variància i la desviació típica:

Analizar / Estadísticos descriptivos / Descriptivos / Opciones

I ara seleccionem la mitjana, la variància i la desviació típica. Obtenim el resultat següent:

Estadísticos descriptivos

Región económica		N	Media	Desv. típ.	Varianza
OCDE	Producto interior bruto per-capita	21	16610,86	3725,97	1,4E+07
	N válido (según lista)	21			
Europa Oriental	Producto interior bruto per-capita	14	5159,79	1708,70	2919658
	N válido (según lista)	14			
Asia / Pacífico	Producto interior bruto per-capita	17	4263,00	6291,05	4,0E+07
	N válido (según lista)	17			
África	Producto interior bruto per-capita	19	998,68	1178,26	1388291
	N válido (según lista)	19			
Oriente Medio	Producto interior bruto per-capita	17	4957,41	4057,45	1,6E+07
	N válido (según lista)	17			
América Latina	Producto interior bruto per-capita	21	1997,67	1482,12	2196689
	N válido (según lista)	21			

Observem que ha realitzat el càlcul com si es tractés de 6 mostres diferents, però ens presenta la taula de forma que és fàcil comparar els resultats dels descriptius del producte interior brut per a cada un dels grups que ens defineix la variable **region**.

► Exercici 4

Amb la mateixa segmentació per regió, demaneu un diagrama de sectors per a la variable religió: què obtenim?

És **molt important** que un cop haguem realitzat l'estudi que volíem fer ens recordem de **desfer la segmentació** per tornar a treballar amb l'arxiu complet, com una sola mostra. Per fer-ho, el menú és

Datos / Segmentar archivo / Analizar todos los casos, no crear grupos

- **Organitzar els resultats per grups.** L'altra opció quan necessitem segmentar un arxiu de dades és organitzar els resultats per grups. El menú en aquest cas és:

Datos / Segmentar archivo / Organizar los resultados por grupos

escollim la variable **region** i premem acceptar.

Si ara tornem a fer el càlcul dels descriptius per a la variable **pib_cap** veurem que torna a fer el mateix càlcul però ens presenta els resultats de diferent forma:

Región económica = OCDE

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	21	16610,86	3725,97	1,4E+07
N válido (según lista)	21			

a. Región económica = OCDE

Región económica = Europa Oriental

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	14	5159,79	1708,70	2919658
N válido (según lista)	14			

a. Región económica = Europa Oriental

Región económica = Asia / Pacífico

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	17	4263,00	6291,05	4,0E+07
N válido (según lista)	17			

a. Región económica = Asia / Pacífico

Región económica = África

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	19	998,68	1178,26	1388291
N válido (según lista)	19			

a. Región económica = África

Región económica = Oriente Medio

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	17	4957,41	4057,45	1,6E+07
N válido (según lista)	17			

a. Región económica = Oriente Medio

Región económica = América Latina

Estadísticos descriptivos^a

	N	Media	Desv. típ.	Varianza
Producto interior bruto per-capita	21	1997,67	1482,12	2196689
N válido (según lista)	21			

a. Región económica = América Latina

Recordeu que un cop heu acabat l'estudi segmentat cal **desfer la segmentació** de l'arxiu.

► Exercici 5

Seguim amb l'arxiu **Mundo 95.sav**; volem estudiar la distribució de països en regions econòmiques en dos grups, els que tenen producte interior brut per càpita superior a 10 000 i els que el tenen inferior o igual.

Amb les eines que coneixem fins ara, podem fer-ho de la següent forma:

- Seleccionem els individus que satisfan una d'aquestes dues condicions. En fer la selecció, l'SPSS es crea una variable de filtratge que s'anomena **filter_\$**. Li canviem el nom per **filtre_pib**.
- Desfem la selecció, però ens ha quedat la variable filtre_pib que pren valors 1 o 0 segons si l'individu (en aquest cas el país) satisfà o no la condició anterior.
- Segmentem l'arxiu segons els valors d'aquesta variable de filtratge filtre_pib.
- Finalment podem fer, per exemple, un diagrama de sectors per a la variable region. Interpreteu el resultat.